이산 웨이브렛 변환영역에서의 스펙트럼 차감법을 이용한 잡음제거

김현기[†] · 이상운[††] · 홍재근[†††]

요 약

잡음환경에서의 음성인식을 위하여 음성에 부가된 잡음을 제거하는 방법에 있어, 기존의 스펙트럼 차감법은 잡음과 음성을 정확히 구별하기 힘들고 정확한 잡음의 특성을 추정할 수 없는 단점이 있다. 또한, 웨이브렛 변환영역에서의 잡음제거 방법은 임계값 적용시 저주파 영역보다는 고주파영역에 상대적으로 더 큰 영향을 미쳐 고주파영역에서 신호의 손실이 발생하는 단점이 있다.

본 논문에서는 스펙트럼 차감법 및 웨이브렛 변환을 이용한 잡음제거 방법의 단점을 개선하기 위하여 연속 웨이브렛 변환 영역에서 웨이브렛 계수의 스케일별 표준편차로 묵음구간과 음성구간을 판별하여 끝점을 검출 후, 잡음이 섞인 음성신호를 이산 웨이브렛 변화에 의해 3개의 대역으로 분리하여 각각의 대역 내에서 스펙트럼 차감법을 적용시키는 방법을 제안한다. 끝점을 검출하고 대역을 나눔으로써 스펙트럼 차감을 적용할 잡음 신호의 특성을 정확히 추출할 수 있다. 실험을 통하여 제안한 방법이 기존의 스펙트럼 차감법 및 웨이브렛 변환을 이용한 잡음제거 방법보다 신호대 잡음비 및 Itakura-Saito 거리 측면에서 향상됨을 확인할 수 있었다.

Noise Reduction using Spectral Subtraction in the Discrete Wavelet Transform Domain

Hyun Ki Kim[†], Sang Woon Lee[††] and Jae Keun Hong[†††]

ABSTRACT

In noise reduction method from noisy speech for speech recognition in noisy environments, conventional spectral subtraction method has a disadvantage which distinction of noise and speech is difficult, and characteristic of noise can't be estimated accurately. Also, noise reduction method in the wavelet transform domain has a disadvantage which loss of signal is generated in the high frequency domain. In order to compensate these disadvantage, this paper propose spectral subtraction method in continuous wavelet transform domain which speech and non-speech intervals is distinguished by standard deviation of wavelet coefficient, and signal is divided three scales at different scale. The proposed method extract accurately characteristic of noise in order to apply spectral subtraction method by end detection and band division. The proposed method shows better performance than noise reduction method using conventional spectral subtraction and wavelet transform from viewpoint signal to noise ratio and Itakura-Saito distance by experimental.

1. 서 론

디지털 신호처리 기술과 컴퓨터 및 통신기술의 발달로 인간과 기계간의 의사소통을 보다 자연스럽고 정확하게 하고자 하는 man-machine interface 기술이 현실적 문제로 부각되고 있는 가운데 최근에는 멀티미디어나 이동통신과 같은 다양한 정보매체를 통한 통신분야에 있어서도 음성신호처리의 중요성이 대두되고 있다. 특히 음성인식 시스템의 실용화가

[†] 정회원, 경남정보대학 전자정보학부 교수
[††] 포항1대학 컴퓨터응용과 전임강사
[†††] 경북대학교 공과대학 전자공학과 교수

늘어남에 따라 최근에는 주변 잡음에 대한 인식 시스템의 성능저하가 문제시되고 있다. 그러한 이유는 잡음이 없거나 비교적 조용한 실험실 환경에서는 우수한 성능을 나타내는 음성인식 시스템의 성능이 잡음이 존재하는 곳에서는 급격히 떨어지기 때문이다. 실제로 사무실이나 실험실이 아닌 상황에서 발생하는 잡음의 영향으로 생기는 문제점은 화자의 발음이 조용한 환경에서 발음한 것과 다르고, 음석인식 시스템의 입력으로 잡음과 음성신호가 동시에 들어오기 때문에 기준패턴과 다른 형태를 갖는다는 것이다. 따라서 음성에 부가된 잡음을 제거하고자 하는 연구와 잡음이 부가된 음성에서 잡음에 강한 특징 파라미터의 추출에 관한 연구가 오랫동안 수행되어 왔다[1].

일반적으로 음성에 포함된 잡음을 제거하는 방법은 주파수영역에서 음성신호와 음성신호에 섞인 잡음을 분리시켜서 잡음에 의한 성분을 제거시키는 것이다. 왜냐하면 시간영역에서 음성신호와 그에 섞인 잡음을 구분하기가 힘들기 때문이다. 스펙트럼 차감법(spectral subtraction method)[2]은 음성신호와 잡음신호가 주파수 영역에 부가적이라고 가정하고, 이산 푸리에 변환(discrete Fourier transform)을 통하여 잡음에 의한 주파수 성분의 크기를 잡음이 섞인 음성신호의 주파수 성분의 크기에서 차감하는 것이다. 차감된 주파수 성분은 역푸리에 변환에 의해 잡음이 제거된 깨끗한 음성신호로 복원된다. 이 방법에 사용된 이산 푸리에 변환은 음성신호가 정상적(stationary)이라는 가정에서 적용된다. 음성신호는 비정상적(nonstaionary)인 특성을 가진다. 단구간 푸리에 변환(short-time Fourier)은 음성신호를 짧은 구간으로 나누어 푸리에 변환을 적용한다. 일반적으로 짧은 구간으로 나누는 방법은 창함수를 이용한다. 창함수내의 음성신호는 거의 정상적인 특성을 가지게 된다. 그러나, 창함수의 길이에 의해 주파수 영역의 분해능은 결정된다. 또한 고정된 창함수의 길이에 의한 일정한 분해능은 해석하고자 하는 신호의 특성에 따라 창함수의 길이를 가변시켜야 하는 단점과 연관된다.

최근 들어 활발히 연구되고 있는 웨이브렛 변환(wavelet transform)은 해석하고자 하는 주파수 성분에 따라 가변할 수 있는 창함수의 크기를 제공한다. 웨이브렛 변환에서 창함수는 기저함수(basis function)라고 불리어진다. 즉 가변하는 기저함수에 의한 웨이브렛 변환의 다중해상도 특성과 시간-주파수 국부성(time-frequency localization)은 통계적 특성을 모르거나 시간적으로 예측하기 힘든 비정상(nonstationary)상태의 신호해석에 매우 유용한 것으로 밝혀졌다[3]. 웨이브렛 변환에 의한 잡음 제거 방법은 주로 음성신호가 없는 묵음구간에서 구한 잡음신호의 웨이브렛 변환 계수의 평균 또는 최대치의 일부분을 임계치로 결정한 후, 잡음이 섞인 음성신호의 웨이브렛 변환 계수로부터 차감하는 것이다.

본 논문에서는 기존의 전체 대역에 스펙트럼 차감을 적용하는 대신에 웨이브렛 변환에 의한 대역통과 필터를 거친 신호에 대하여 스펙트럼 차감을 적용한다. 차감되는 잡음신호의 특성은 각 대역에서 추출되게 된다. 시간-주파수 국부성이 뛰어난 웨이브렛 변환의 대역별 잡음 스펙트럼 성분은 기존의 방법으로 얻어진 잡음 스펙트럼 성분보다 잡음의 특성을 더 잘 반영할 수 있다. 제안한 웨이브렛 변환을 통한 스펙트럼 차감법은 기존의 스펙트럼 차감법 및 웨이브렛 변환에 의한 잡음 제거법보다 신호대 잡음비를 더 향상시켰고 잡음이 섞인 음성신호와 복원된 음성신호 사이의 Itakura-Saito 거리도 줄였다.

2. 스펙트럼 차감법을 이용한 잡음제거 방법

스펙트럼 차감법은 잡음이 섞인 음성신호의 스펙트럼에서 잡음신호의 스펙트럼을 빼어줌으로써 깨끗한 음성신호의 스펙트럼을 얻고자 하는 것이 기본원리이다. 그러나 이 방법은 잡음의 스펙트럼 형태를 미리 알고 있거나 잡음의 스펙트럼을 추정하기에 충분한 묵음 구간(약 300ms)이 주어져야 하고, 음성이 존재하는 구간과 잡음만이 존재하는 구간을 검출할 수 있는 방법이 필요하다. 이러한 조건을 만족한다고 가정하면 잡음신호 $n(k)$가 음성신호 $s(k)$에 더해졌을 때 손상된 음성신호 $x(k)$는 다음과 같이 나타낼 수 있다.

$$x(k) = s(k) + n(k) \tag{1}$$

식을 푸리에 변환하면

$$X(\omega) = S(\omega) + N(\omega) \tag{2}$$

여기서

$$X(\omega) = \sum_{k=0}^{N-1} x(k) e^{-j\omega k}$$

$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j\omega k} d\omega$$

와 같이 된다. 추정된 음성신호의 스펙트럼 크기는

$$|\hat{S}(\omega)| = |X(\omega)| - |N(\omega)| \quad (3)$$

와 같이 나타나지만 실제로는 잡음의 특성을 정확히 알 수 없으므로 $|N(\omega)|$대신 잡음 스펙트럼의 평균값인 $\mu(\omega)$를 사용하여 아래와 같이 나타낸다.

$$|\hat{S}(\omega)| = |X(\omega)| - \mu(\omega) \quad (4)$$

여기서 $\mu(\omega)$는

$$\mu(\omega) = E\{|N(\omega)|\} \quad (5)$$

이고 음성이 없는 잡음구간에서 수 프레임의 데이터로부터 취한 샘플평균을 의미한다. 크기정보 이외에 위상정보는 음질향상이나 음성인식 과정에서 중요한 변수가 아니라는 것을 고려하여 손상된 음성의 위상정보 $\theta_x(\omega)$를 추정음성의 위상 $\theta_s(\omega)$로 간주하여 사용한다. 따라서 추정된 음성은 아래와 같이 표현할 수 있다.

$$\begin{aligned}\hat{S}(\omega) &= [|X(\omega)| - \mu(\omega)]e^{j\theta_x(\omega)} \\ &= H(\omega)\,X(\omega)\end{aligned} \quad (6)$$

이다. 여기서 $H(\omega)$는 스펙트럼 차감필터의 전달함수이고

$$H(\omega) = 1 - \frac{\mu(\omega)}{|X(\omega)|} \quad (7)$$

이다. 추정된 음성의 스펙트럼을 역푸리에 변환하여 출력 음성신호를 얻을 수 있다.

그러나 잡음의 특성을 정확히 알 수 없으므로 잡음 스펙트럼 $|N(\omega)|$ 대신 평균값인 $\mu(\omega)$를 사용함에 따라 잡음제거 과정에서 다음과 같은 주파수오차(spectral error)가 발생한다.

$$\begin{aligned}\varepsilon(\omega) &= \hat{S}(\omega) - S(\omega) \\ &= N(\omega) - \mu(\omega)\,e^{j\theta_x}\end{aligned} \quad (8)$$

이러한 주파수오차를 줄이는 방법에는 프레임별 크기를 평균하는 방법과 반파 정류 방법이 있다. 스펙트럼 차감법의 잡음제거 과정은 그림 1과 같다. 그림 1에서 $|\cdot|$는 크기를 구하는 것이다.

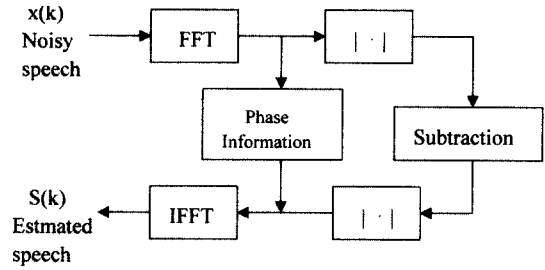



그림 1. 스펙트럼 차감법의 블록도

3. 웨이브렛 변환을 이용한 잡음제거 방법

3.1 웨이브렛 변환

웨이브렛 해석은 웨이브렛이라고 불리는 하나의 원형 함수와 이 함수의 스케일된 함수가 기저함수를 이루게 되며 대역통과필터와 같은 역할을 수행하게 된다. 기저함수는 원형 함수의 천이와 스케일(수축/확장)에 의해 얻어진다.

$$h_{a,b} = \frac{1}{\sqrt{a}}\,h\left(\frac{t-b}{a}\right) \quad (9)$$

여기서 $a \in R^+$, $b \in R$이다. a는 스케일 파라미터이며 b는 천이 파라미터이고 $1/\sqrt{a}$는 정규화를 위해서이다. a가 큰 값일 때 기저함수는 원형함수의 확장된 형태이며 저주파 함수이다. a가 작은 값일 때 기저함수는 원형함수의 수축된 형태이며 고주파 함수이다. 웨이브렛 해석은 웨이브렛 함수가 스케일 파라미터 a에 따라 확장 및 수축함에 의해 시간적-국부성 해상도를 가지게 된다. 웨이브렛 해석에 사용되는 웨이브렛 함수는 다양하게 존재하며 주로 해석하고자 하는 신호의 특성에 맞게 선택되어진다[4-6]. 또는 웨이브렛 해석 과정에서 측도를 사용하기도 한다[7,8]. 음성신호 해석분야는 음성파형과 유사한 모양을 가지는 다우베치(daubechies) 웨이브렛 함수를 주로 사용한다[9].

연속 웨이브렛 변환은 다음과 같이 정의된다.

$$X_W(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} h^*\left(\frac{t-b}{a}\right) x(t)\,dt \quad (10)$$

연속 웨이브렛 변환은 연속적인 스케일 및 천이 파라미터에 의해 계산량이 복잡할 뿐만 아니라 지나친 정보의 량을 가지게 된다.

이에 식(9)와 식(10)에서 천이 및 확장, 수축 파라

미터를 이산화하면

$$h_{mn}(t) = a_0^{-m/2} \cdot h(a_0^{-m}t - nb_0), \quad m, n \in Z, \quad a_0 > 1, \quad b_0 \neq 0 \tag{11}$$

$$X_W(m,n) = a_0^{-m/2}\int_{-\infty}^{\infty} h(a_0^{-m}t - nb_0)x(t)dt \tag{12}$$

이다. 이것은 $a = a_0^m$, $b = n a_0^m b_0$에 대응된다. 특히, $a_0 = 2$, $b_0 = 1$로 이원화하면 식 (13)과 같다.

$$h_{mn}(t) = a_0^{-m/2} \cdot h(a_0^{-m}t - nb_0), \quad m, n \in Z, \quad a_0 = 2, \quad b_0 = 1 \tag{13}$$

그림 2는 신호를 옥타브 대역으로 분할하는 이산 웨이브렛 변환을 나타낸다. $h_0(n)$은 저역통과필터이고 $h_1(n)$은 고역통과필터이며 2↓는 2배로 decimation하는 것이다. 원신호를 $h_0(n)$, $h_1(n)$에 통과시키는 것은 두 대역의 신호로 나누는 것과 같고 저역통과된 신호를 반복적으로 $h_0(n)$, $h_1(n)$에 통과시키면 일정한 Q값을 갖는 옥타브 대역을 형성하게 된다.

3.2 웨이브렛 변환을 이용한 잡음제거

웨이브렛 변환하면 낮은 주파수 대역내의 계수들이 신호의 전체적인 성분을 포함하고 있으므로 높은 주파수 대역을 제거한다 하더라도 원신호를 어느 정도는 복구할 수 있으므로 잡음성분을 웨이브렛 계수들의 임계치이하 되는 성분을 제거하는 방법이 웨이브렛 변환을 이용하여 잡음을 제거하는 방법의 기본이다[10-12].

원신호 x가 평균값이 영이고 표준편차가 σ인 백색 가우시안 잡음(white Gaussian noise)에 의해 손상되었다면 잡음에 손상된 신호 y는 다음과 같이 표현할 수 있다.

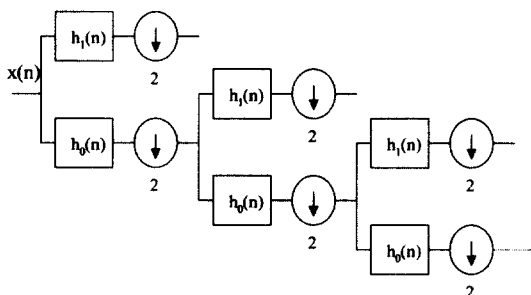



그림 2. 이산 웨이브렛 변환

$$y = x + \sigma n \tag{14}$$

식(14)를 웨이브렛 변환시키면 식(15)와 같이 나타낼 수 있다.

$$Y = X + N \quad (15)$$

나타내고 식(15)에서 여기에서, 대문자는 웨이브렛 영역에서의 신호를 웨이브렛 변환 행렬이 W라면 Y, X,N는 다음과 같다. 즉, $Y = Wy$, $X = Wx$, $N = Wn$이다.

웨이브렛 변환행렬의 역함수를 M이라 표현하면 복구된 원신호 $\hat{x}$는 다음과 같이 표현 할 수 있다.

$$\hat{x} = M\hat{X} = M\hat{Y} \tag{16}$$

여기서 $\hat{Y}$는 특정 임계치 이하를 제거시킨 웨이브렛 계수를 나타내고 $\hat{X}$는 원신호에서 단순히 개개의 웨이브렛 계수를 없애거나 유지함으로써 추정될 수 있다. 위와 같은 가정은 낮은 주파수 대역의 웨이브렛 계수가 원신호의 정보의 대부분을 가지고 있으므로 높은 주파수 대역의 성분을 없애거나 특정 임계값 이하의 값을 제거하더라도 원신호와 가까운 값을 얻을 수 있기 때문이다. 본 논문에서는 식(17)과 같은 soft threshold 방법[13,14]으로 웨이브렛 계수들을 제거하는 방법을 사용하였다.

$$y = \begin{cases} sign(x)\,(|x| - th), & |x| > th \\ 0, & |x| < th \end{cases} \tag{17}$$

웨이브렛 변환을 이용하여 잡음에 손상된 음성에서 잡음이 제거된 음성을 복구하는 과정을 요약하면 다음과 같다.

(1) Decompose

기저 웨이브렛을 선택한 후 잡음에 손상된 음성으로부터 웨이브렛 변환을 취하여 웨이브렛 계수를 구한다.

(2) Threshold coefficients

각각의 레벨별로 임계값을 정한 후 soft threshold 방법을 이용하여 웨이브렛 계수의 임계값이하 성분을 제거한다.

(3) Reconstruction

추정된 웨이브렛 계수로부터 역 웨이브렛 변환을

취함으로써 잡음이 제거된 음성을 복구한다.

4. 이산 웨이브렛 변환 영역에서 스펙트럼 차감법을 이용한 잡음제거 방법

우선 잡음신호의 특성을 알아내기 위해 음성신호가 있는 구간과 잡음 신호만이 있는 구간을 구별할 필요가 있다. 이러한 구별과정을 끝점검출이라고 한다. 끝점 검출이 수행된 후 잡음 신호만이 존재하는 구간에 서 잡음의 특성을 추출하므로 끝점 검출은 잡음신호 제거에 선행되어야 한다. 끝점 검출법은 단구간 에너지를 이용하는 방법과 영교차율을 이용하는 방법이 있다. 이러한 방법은 구현은 간단하지만 잡음이 섞인 신호에 대해서는 제대로 수행되지 않는다. 본 실험에 사용된 끝점 검출 방법은 연속 웨이브렛 변환으로 얻어지는 웨이브렛 계수가 잡음 신호 구간과 잡음이 섞인 음성신호 구간에서 다른 특성을 가진다는 것을 이용한다. 웨이브렛 계수의 표준편차는 잡음 신호 구간에서 작은 값을 가지며 음성이 포함된 구간에서는 비교적 큰 값을 가진다. 연속 웨이브렛 변환에서 스케일 값이 작을수록 빠르게 변화하는 부분을 정확하게 검출할 수 있어서 고주파 구간에서 웨이브렛 계수의 표준편차는 작아지게 된다. 잡음이 섞인 음성신호의 초기부분은 잡음신호만이 존재한다고 가정한다. 끝점은 초기 잡음 신호 구간의 웨이브렛 계수의 표준편차의 평균값에 일정한 상수값을 곱하여 임계값으로 설정하고 이 임계값보다 큰 표준편차를 가지는 구간은 음성이 존재하는 것으로 판별하여 검출된다. 이 때 곱해지는 상수값은 잡음정도에 따라 실험적으로 선정한다. 임계값 T_H는 다음 수식에 따라 구할 수 있다.

$$T_H = c \cdot E[sd_i] \tag{18}$$

여기서, c는 상수값이고 sd_i는 i스케일 웨이브렛 계수의 표준편차를 나타내고 $E(\cdot)$는 평균을 취하는 것이다. 또한, 표준편차의 변별력은 연속 웨이브렛 변환에 의한 대역별 웨이브렛 계수에 따라 달라진다. 적절한 대역의 웨이브렛 계수는 자음과 모음의 주파수 성분이 다르다는 것을 이용하여 선택한다.

끝점 검출에 의해 구해진 잡음이 섞인 음성은 스펙트럼 차감법에 적용되게 된다. 스펙트럼 차감은 이산 웨이브렛 변환에 의해 전 대역을 몇 개의 대역으로 나누어 각 대역별로 수행된다. 따라서 잡음신호의 특성은 각 대역별로 구할 수 있어 전 대역에서 구한 것보다 좀 더 잘 표현될 수 있다. 대역의 수는 스펙트럼 차감법을 수행할 계산량과 각 대역에서 원신호를 어느 정도 나타낼 수 있는가에 따라 결정한다. 전체 알고리듬을 그림 3에 나타내었다.

스펙트럼 차감법을 적용한 후에도 여전히 묵음 구간에는 잡음성분이 존재하게 된다. 또한 스펙트럼 차감 과정에서 추정한 음성신호에 주파수 오차로 인해 음악 잡음이 존재하게 된다. 스펙트럼 차감 적용 부분에 몇가지 부가적인 기능을 첨가하여야 한다. 잡음신호 억압과정과 잔여(residual) 잡음 제거 과정이 있다.

잡음신호 억압과정은 끝점 검출에 의해서 구해진 묵음구간에 대해 각 대역별로 프레임별 전력 $\hat{S}(w)$과 평균적인 잡음 전력 $\mu(w)$의 비가 임계값 T_H보다 작은 경우 감쇠 상수를 적용한다. 즉, 추정된 음성신호의 주파수 성분 $\hat{S}(w)$는

$$\hat{S}(w) = C\,\hat{S}(w) \quad \text{if } |\hat{S}(w)| < T_H \cdot \mu(w) \tag{19}$$

가 된다. 여기서, C는 감쇠상수이다.

잔여 잡음 제거 과정은 추정된 신호에 대해서 묵음구간에서 최대 잡음 스펙트럼 크기 값을 정한 후, 모든 프레임에 대한 스펙트럼의 크기의 비 λ를 구한다.

$$\lambda = \frac{|\hat{S}(w)|}{\max|\hat{N}(w)|} \tag{20}$$

λ가 1보다 클 경우는 음성신호가 있는 구간으로 간주하고 λ가 1보다 작을 경우에는 잡음이 있는 구

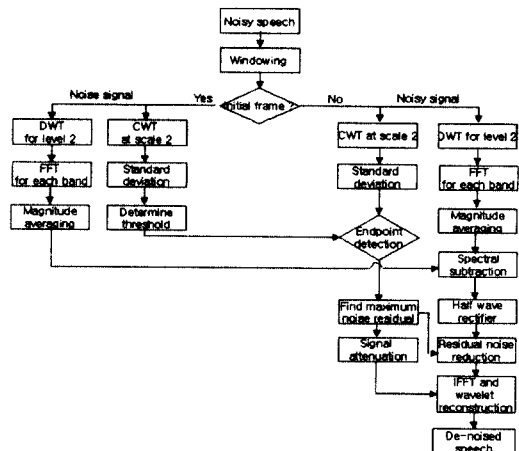



그림 3. 웨이브렛 변환을 이용한 스펙트럼 차감법

간으로 간주한다. 모든 프레임의 스펙트럼 크기는 인접한 세 개의 프레임에서 가장 작은 스펙트럼 크기를 지정한다.

5. 실험결과 및 고찰

원음성은 "대한민국"을 발음한 데이터이고 16K Hz의 표본주파수를 가진다. 원음성에 백색 가우시안 잡음을 첨가한 후, 프레임은 256개의 샘플로 형성하고 중첩은 128샘플로 하였다. 먼저, 끝점 검출은 스케일 값이 2인 연속 웨이브렛 변환 계수의 표준편차 평균을 이용한다. 이 때 연속 웨이브렛 변환에 사용되는 웨이브렛 기저함수는 시간적으로 분해능이 뛰어난 Mexican hat[15] 함수를 사용하였다. 잡음이 섞인 음성신호의 초기 구간(300ms)은 잡음 성분만이 존재하므로 이 구간에서 임계값을 결정한다. 프레임의 표준편차가 임계값보다 연속하여 5프레임 이상 클 경우에 음성이 존재하는 프레임으로 판별하고 마찬가지로 연속적으로 5프레임 이상 임계값보다 작은 경우에는 잡음만이 존재하는 프레임으로 판별하여 끝점 검출을 하게 된다.

Mexican hat 웨이브렛 함수에 의한 연속 웨이브렛 변환 계수의 프레임별 표준편차를 스케일 값에 따라 그림 4와 그림 5에 나타내었다.

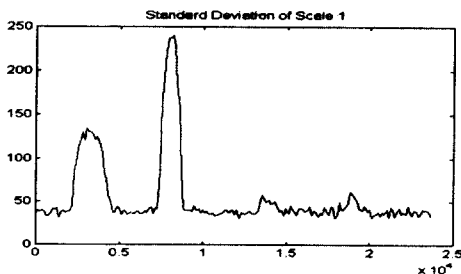



그림 4. 스케일 1의 표준편차

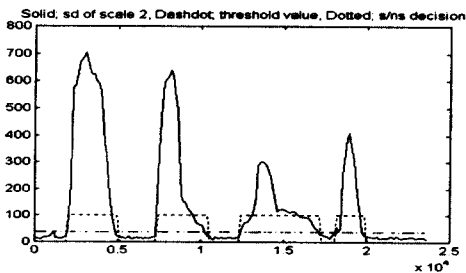



그림 5. 스케일 2의 표준편차 및 임계치

그림 4는 스케일 값이 1일 때, 그림 5는 스케일 값이 2일 때의 프레임별 표준편차를 나타낸다. 두 그림에서 고주파 성분이 강한 잡음과 자음부분은 작은 값을 가지며 모음부분에서는 큰 값을 가짐을 확인할 수 있다. 스케일에 따라 프레임별 표준편차는 다른 값을 나타내고 있다. 스케일 값이 1인 경우에는 잡음신호 구간뿐만 아니라 일부 모음 구간 /i/,/u/에서도 작은 값의 표준 편차를 가진다. 스케일 값이 2인 경우에는 웨이브렛 함수가 확장(dilation)되어 상대적으로 높은 주파수를 가지는 /i/, /u/의 모음 구간에서 표준편차의 값을 커지게 한다.

잡음 신호를 제거하는 방법을 비교하기 위해서 기존의 스펙트럼 차감법과 웨이브렛 변환에 의한 방법과 제안한 웨이브렛 변환에 의한 스펙트럼 차감법을 실험하였다. 제안한 방법에서 이산 웨이브렛 변환은 세 개의 대역을 가지도록 분해(decomposition)하였으며 웨이브렛 함수는 4개의 탭을 가지는 다우베치(Daubechies)함수를 사용하였다. 다우베치 웨이브렛 함수를 사용한 이유는 함수 모양이 음성 파형과 유사하고 직교성을 가지기 때문이다. 직교성을 가지는 웨이브렛 함수는 원신호를 완전히 복원할 수 있다. 다우베치 웨이브렛에서 탭의 길이는 필터 계수의 개수와 동일하며 탭의 길이가 커질수록 대역통과 필터와 같은 역할을 수행한다. 4개의 탭을 가지는 다우베치 웨이브렛 함수는 필터 계수가 작아서 웨이브렛 변환 계산량을 줄일 뿐만 아니라 대역통과필터와 같은 특성을 가진다. 각 방법의 결과를 아래의 그림에 나타내었다. 그림 6은 원신호를 나타내고 그림 7은 10dB의 신호대 잡음비를 가지도록 백색 잡음을 원신호에 추가한 신호를 나타낸다.

그림 8, 그림 9 및 그림 10에 각각 손상된 음성에 대해 기존의 스펙트럼 차감법 및 웨이브렛 변환으로

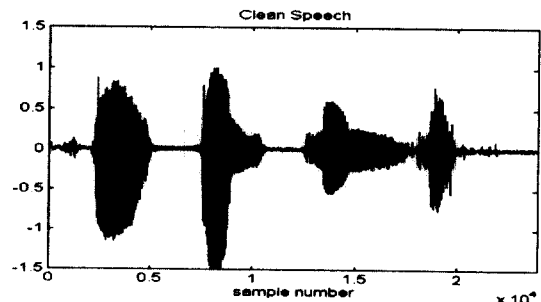



그림 6. 원 음성신호

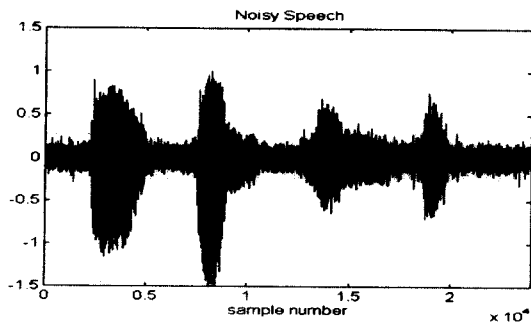



그림 7. SNR이 10dB인 잡음섞인 음성신호

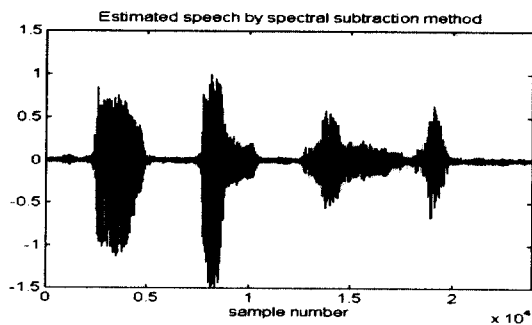



그림 8. 기존의 스펙트럼 차감법으로 추정한 음성신호

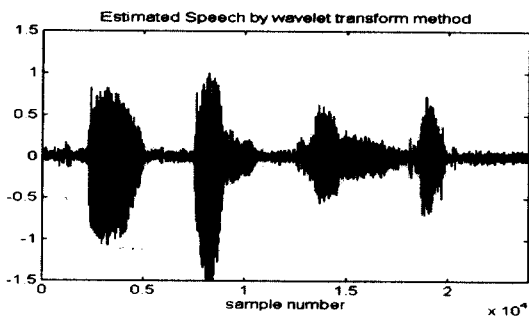



그림 9. 웨이브렛 변환법으로 추정한 음성신호

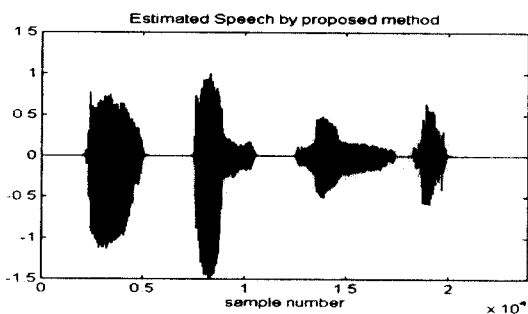



그림 10. 제안한 방법으로 추정한 음성신호

추정된 음성신호와 제안한 방법으로 추정된 음성신호를 나타내었다. 파형신호의 고찰에 의해 웨이브렛 변환법은 여전히 묵음구간에서 잡음 성분이 존재한다는 것을 알 수 있다.

그림 11은 신호대 잡음비가 5dB인 비교적 큰 잡음이 첨가된 손상된 음성신호를 나타내고 있다.

그림 12, 그림 13 및 그림 14에 각각 이 손상된 음성에 대해 기존의 스펙트럼 차감법 및 웨이브렛 변환으로 추정된 음성신호와 제안한 방법으로 추정된 음성신호를 나타내었다.

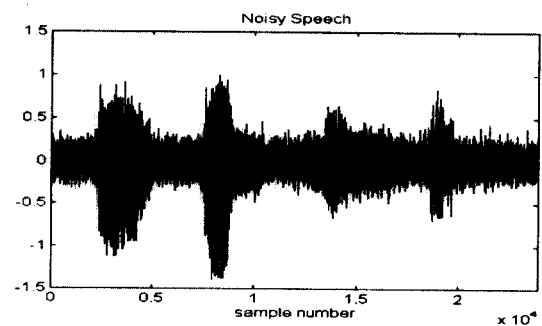



그림 11. SNR이 5dB인 잡음섞인 음성신호

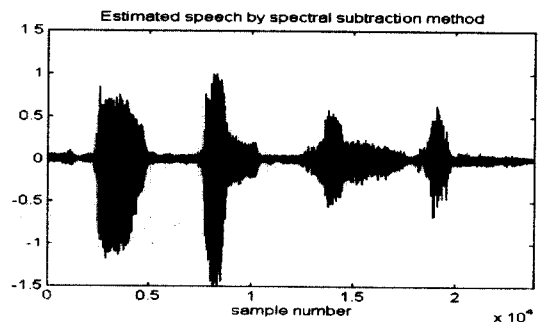



그림 12. 기존의 스펙트럼 차감법으로 추정한음성신호

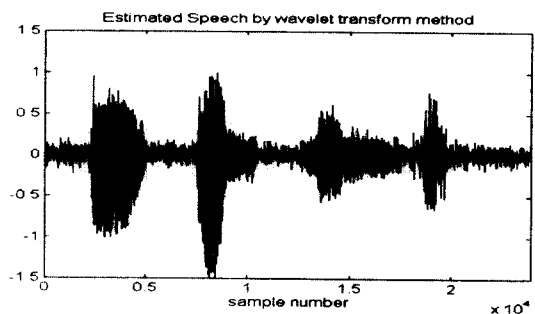



그림 13. 웨이브렛 변환법으로 추정한 음성신호

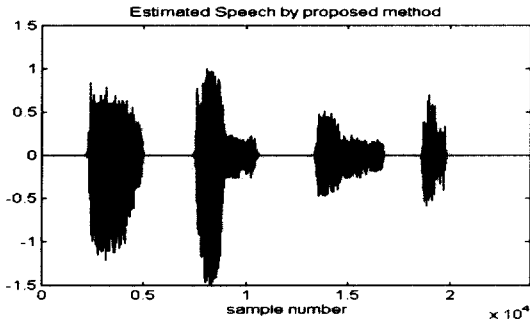



그림 14. 제안한 방법으로 추정한 음성신호

기존의 방법으로 추정한 음성신호와 제안한 방법으로 추정한 음성신호는 신호대 잡음비가 비교적 큰 경우에는 별로 차이가 나지 않으나 신호대 잡음비가 적은 경우에는 큰 차이가 있음을 볼 수 있다.

기존의 스펙트럼 차감법, 웨이브렛 변환법 및 제안한 방법에서 추정된 음성신호의 신호대 잡음비를 표 1에 나타내었다. 제안한 방법이 기존의 스펙트럼 차감법 및 웨이브렛 변환을 이용한 잡음제거 방법에 비해 신호대 잡음비의 향상 정도가 크고 잡음이 많이 섞인 음성 신호에서 개선되는 정도가 크다는 것을 알 수 있고 유사도를 나타내는 Itakura-Saito (IS) 거리도 감소되어 원음에 가깝다는 것을 알 수 있었다. 또한, 스펙트럼 차감법에서 SNR이 10dB인 경우에 가장 적은 IS 거리를 나타내는 것은 원음(clean speech)이 완전히 잡음이 없는 상태에서 녹음된 것이 아니라 10dB에서 15dB사이의 신호임을 추정할 수 있다.

그림 15는 원 음성신호와 추정된 음성신호의 LPC

표 1. 추정된 음성신호의 신호대 잡음비(dB)와 IS 거리 (IS_SN: 원신호와 잡음신호간의 IS 거리, IS_SR: 원신호와 추정된 음성신호간의 IS 거리)

SNR[dB]		20	15	10	5	0
스펙트럼 차감법	SNR	26.73	23.83	21.83	18.96	13.93
	IS_SN	0.46	0.82	1.30	1.81	2.31
	IS_SR	9.99	2.64	0.43	0.65	1.10
웨이브렛 변환법	SNR	20.55	18.77	17.90	14.36	10.08
	IS_SN	0.46	0.82	1.30	1.81	2.31
	IS_SR	0.07	0.37	0.51	0.63	3.41
제안한 방법	SNR	34.63	30.41	26.13	22.46	19.06
	IS_SN	0.46	0.82	1.30	1.81	2.31
	IS_SR	0.02	0.05	0.25	0.45	0.83

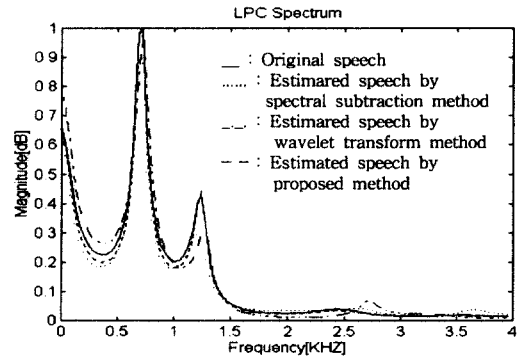



그림 15. LPC 스펙트럼 (실선 : 원 음성신호, 점선: 기존의 스펙트럼 차감법으로 추정한 음성신호, 실/점선 : 웨이브렛 변환 방법으로 추정한 음성신호, 파선 : 제안한 방법으로 추정한 음성신호)

스펙트럼 특성을 나타내고 있다. 제안한 방법에 의한 LPC가 원 음성신호의 하곡부를 따르고 포먼트를 더 잘 나타내고 있다.

6. 결 론

본 논문에서는 음성에 부가된 잡음을 제거하는 방법에 있어, 스펙트럼 차감법에서 나타나는 잡음과 음성의 구별이 어려운점과 정확한 잡음의 특성을 추정할 수 없는 단점을 극복하고, 웨이브렛 변환법에서 나타나는 고주파 영역에서의 신호의 손실을 감소시키기 위하여, 이산 웨이브렛 변환 영역에서 스펙트럼 차감법을 이용하여 잡음을 제거시키는 방법을 제안하였다.

제안한 방법에서는 원 신호를 3개의 대역을 가지도록 분해하여 각각의 대역에서 신호의 특성을 나타내는 웨이브렛 계수를 웨이브렛 변환을 통하여 얻은 후 이 웨이브렛 계수를 스펙트럼 차감법에 적용시켰다. 웨이브렛 변환과정에서 부가적인 잡음제거 효과와 각각의 스케일에서 잡음 스펙트럼 특성에 좀 더 근사하는 평균값을 구할 수 있어서 기존의 스펙트럼 차감법 및 웨이브렛 변환 방법보다 신호대 잡음비와 IS 거리 측면에서 향상되었다. 그리고 제안한 방법으로 추정한 음성의 주파수 특성이 기존의 방법으로 추정한 음성의 주파수 특성에 비해 원 음성에 근접함을 확인하였다. 또한 실제로 추정한 음성을 들려주었을 때 음악잡음이 감소되어 음질이 향상됨을 확인하였다.

참 고 문 헌

[1] L. R. Rabiner, *Fundamentals of speech recognition*, Prentice-Hall, 1993.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. ASSP*, vol. ASSP-27, no. 2, Apr. 1979.

[3] Ingrid Daubechies, "The Wavelet transform, time-frequency localization and signal. analysis," *IEEE Trans. on information theory*, vol. 36, no. 5, Sep. 1990.

[4] Yann Soon, Soo Ngee Koh, and Chai kiat Yeo, "Wavelet for speech denoising," *IEEE TENCON Speech and Image Tech. for Computing and Telecommunication*, pp. 479-482, 1997.

[5] Jing Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis", *NDT & E International*, vol. 34, pp. 25-30, January 2001.

[6] N. S. D. Brito, B. A. Souza, and F. A. C. Pires, "Daubechies wavelets in quality of electrical power," *8th International Conference On Harmonics and Quality of Power Proceedings*, vol 1, pp. 511-515, 1998.

[7] P. Bodin and L. F. Villemoes, "Spectral subtraction in the time-frequency domain using wavelet packet," *IEEE*, 1997.

[8] E. Lukasik, "Wavelet packets based features selection for voiceless plosives classification," *IEEE ICASP*, vol. 2, pp. 689-692, 2000.

[9] J. I. Agbinya, "Discrete wavelet transform techniques in speech processing," *IEEE TENCON. Digital Signal Processing Applications*, vol. 2, pp. 514-519, 1996.

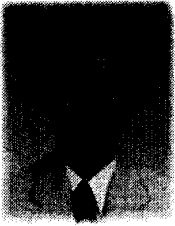
[10] Martin Vetterli and Cormac Herley, "Wavelet and filter banks: Theory and design," *IEEE Trans. Signal Proc.*, vol. 40, no. 9, Sep. 1992.

[11] O. Rioul and M. Vetterli, "Wavelet and signal processing," *IEEE Signal Processing Magazine*, pp. 14-38, Oct. 1991.

[12] Christian van den Branden Lambrecht, "Wavelet packets-based high-resolution spectral estimation," *Signal processing 47*, pp. 135-144, 1995.

[13] D. L. Donoho, "De-Noising by soft thresholding," *IEEE Trans. on Info. Theory*, pp. 613-627, May 1995.

[14] A. Teolis and J. J. Benedetto, "Noise suppression using a wavelet model," *IEEE Trans. ASSP*, vol. I, pp. 17-20, 1994.

[15] P. Palisson, N. Zegadi, F. Peyrin, and R. Unterreiner, "Unsupervised texture segmentation using wavelet decomposition", *IEEE*, pp. 625-627, 1994.

김 현 기

1986년 2월 경북대학교 공과대학 전자공학과 졸업(공학사)
1988년 2월 경북대학교 대학원 전자공학과 졸업(공학석사)
2000년 2월 경북대학교 대학원 전자공학과 졸업(공학박사)
1988년 2월 - 1995년 8월 한국전자통신연구원 미디어연구실 선임연구원
1993년 4월~1994년 5월 미국 Bicom Product Inc. 객원 연구원
1995년 9월~현재 경남정보대학 전자정보학부 교수
관심분야 : 멀티미디어 시스템, 음성인식, 음성신호 처리

이 상 운

1997년 경북대학교 공과대학 전자공학과 공학사
1999년 경북대학교 대학원 전자공학과 공학석사
2001년 경북대학교 대학원 전자공학과 박사수료
2000년 3월~현재 포항1대학 컴퓨터응용과 전임강사
관심분야 : 음성인식, 음성신호 처리

홍 재 근

1975년 경북대학교 전자공학과 공학사
1979년 경북대학교 전자공학과 공학석사
1985년 경북대학교 전자공학과 공학박사
1979~1982 경북산업대학 전임강사, 조교수
1983~현재 경북대학교 공과대학 전자공학과 교수
관심분야 : 음성인식, 음성코딩과 합성, 음질향상, 신호처리시스템개발